Europäisches Patentamt	European Patent Office	Office européen des brevets

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

00480011.6

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

I.L.C. HATTEN-HECKMAN

DEN HAAG, DEN
THE HAGUE, 13/04/00
LA HAYE, LE

EPA/EPO/OEB Form 1014 - 02.91

Europäisches Patentamt | European Patent Office | Office européen des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.: 00480011.6
Demande n°:

Anmeldetag:
Date of filing: 06/01/00
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
INTERNATIONAL BUSINESS MACHINES CORPORATION
Armonk, NY 10504
UNITED STATES OF AMERICA

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
Enqueuing apparatus for virtual circuit merging in a router of an asynchronous transfer mode (ATM) network

In Anspruch genommene Prioriät(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat: State: Pays:	Tag: Date: Date:	Aktenzeichen: File no. Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:
/

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE
Etats contractants désignés lors du depôt:

Bemerkungen:
Remarks:
Remarques:

ENQUEUING APPARATUS FOR VIRTUAL CIRCUIT MERGING IN A ROUTER OF AN ASYNCHRONOUS TRANSFER MODE (ATM) NETWORK

Technical field

The invention relates to the merging on a single virtual circuit (VC) of connections having the same destination in a switching node of an Asynchronous Transfer Mode (ATM) network and in particular to an enqueuing apparatus for VC merged connections in a router of an ATM Network.

Background

The use of ATM switching nodes in an IP network is now the most attractive solution since ATM hardware switches have been extensively studied and are widely available in many different architectures. In the ATM switching mechanism, the implementation of label swapping requires to link the routers by virtual paths (VP)/virtual circuits (VC).

Traditionally, AAL 5 has been used as the encapsulation method in data communication since it is simple, efficient and has a powerful error detection mechanism. For the ATM switch to forward incoming cells to correct outputs, IP route information needs to be mapped to ATM the labels which are kept in the VP and/or VC fields.

Several methods of mapping the route information to ATM labels exist. In the simplest form, each source destination pair is mapped to a unique VC value at a switch. This method, called the non-VC merging case, allows the receiver to easily reassemble cells into respective packets since the VC values can be used to distinguish the senders. However, if there are n sources and m destinations, each switch is potentially required to manage n.m VC labels for full-meshed connectivity. Therefore, this method is not scaleable to large networks.

In the second method called VP merging, the VP labels of cells that are intended for the same destination are translated to the same outgoing VP value, thereby reducing VP consumption downstream. For each VP, the VC value is used to identify the sender so that the receiver can reconstruct packets even though cells from different packets are allowed to interleave. Although the number of label entries is considerably reduced, VP merging is not practical insofar as the VP space is limited (4.096 entries) at the network-to-network interface.

A third method, called VC merging, maps incoming VC labels for the same destination to the same outgoing VC label. This method is scaleable and does not have the space constraint problem as in VP merging. With VC merging, cells for the same destination are indistinguishable at the output of a switch. Therefore, cells belonging to different packets for the same destination must not interleave with each other, or else the receiver will not be able to reassemble the packets. With VC merging, the

boundary between two adjacent packets are identified by the "End-of-Packet" (EOP) marker used by AAL 5.

If VC merging is implemented, each output adapter of a switching node must contain a number of reassembly buffers, followed by a merging unit and an output buffer. Each reassembly buffer corresponds to an incoming VC value. The purpose of the reassembly buffer is to ensure that cells for a given packet do not interleave with other cells that are merged to the same VC. This mechanism (call store-and-forward at the packet level) can be accomplished by storing each incoming cell for a given packet in the reassembly buffer until the last cell of the packet arrives. When the last cell arrives, all cells in the packet are transferred in an atomic manner to the output buffer for transmission to the next hop. To be efficient, such a process must be performed the most quickly as possible. Furthermore, if the traffic is composed of a single class of service (for example best-effort) there is no problem. But, if the traffic is composed of multiple classes of services, only a partial merging is implemented wherein only Vcs destined for the same class are mapped to the same outgoing VC.

Summary of the invention

Accordingly, a first objet of the invention is to provide a mechanism enabling an efficient VC merging in a switching node of an ATM network and avoiding useless operation.

A second object of the invention is to provide an enqueuing apparatus in the switching node of an ATM network for performing a VC merging of all connections for the same destination whatever the class of service they belong to.

The invention relates therefore to an apparatus for enqueuing cells belonging to a plurality of connections corresponding to several classes of service in a first switching node of an

Asynchronous Transfer Mode (ATM) network wherein at least a set of connections of a same class of service are merged in a single virtual circuit (VC) connecting the first switching node to a second switching node, and comprising a data buffer for storing the cells received by the first switching node, a queuing area composed of at least a set of queues associated respectively with the connections of the set of connections and a scheduled queue corresponding to the class of service, and wherein the contents of the set of queues are transferred into the scheduled queue before being transmitted on the virtual circuit. This apparatus is characterized in that each queue of the set of queues is associated with a Reassembly Queue Control Block (RQCB) defining a chain of Buffer Control Blocks (BCB) wherein each BCB corresponds to a cell belonging to a packet transmitted on the associated connection and is composed of the next buffer address in the data buffer and a lock bit (BCB lock) which is normally set to 1 and set to 0 only if the cell is the last cell of the packet, the chain of BCBs being transferred to a Scheduled Queue Control Block (SQCB) associated with the scheduled queue when it is detected that the lock bit of the cell being stored in the data buffer is set to 0, the corresponding BCB being chained to the chain of BCBs in the SQCB without being previously queued in the RQCB.

Brief description of the drawings

The above and other objects, features and advantages of the invention will be better understood by reading the following more particular description of the invention in conjunction with the accompanying drawings wherein :

- Fig. 1 represents a block diagram of protocol engine of an output adapter in ATM switching node wherein the apparatus according to the invention is implemented.
- Fig. 2 is a schematic block diagram representing the data flow of a cell using the control blocks which are

essential features of the apparatus according to the invention.

- Fig. 3 is a schematic block diagram representing the data flow of the cells in the queuing area using the RQCBs and SQCBs illustrated in Fig. 2.
- Fig. 4 is a flow chart of the reassembly enqueuing process implemented in the apparatus according to the invention.
- Fig. 5 is a flow chart of the aging mechanism implemented in the apparatus according to the invention

Detailed description of the invention

Before describing the enqueuing apparatus according to the invention, the following description in reference to Fig. 1 relates to the structure of the Common ATM Processor (CAP) in the output adapter including such an apparatus and the operations performed therein. Upon reception of a cell from the Common ATM Data Mover (CAD), the cell identification unit 10, here an ATM label decode, determines whether the cell is an ATM user cell and sends it to the lookup unit 12 or a control cell from the Control Point and sends it to the control unit 14.

Lookup unit 12 resolves the ATM label (VP-VC) to identify the connection using Control Blocks 16 set by the Control Point. When the cell belongs to a known connection, it is sent to the Traffic Management and Congestion Unit 18. If not, it is sent to the Xmit Forwarding Information unit 20 for discarding.

Traffic Management and Congestion unit 18 classifies the cell according to its class of service found in Control Blocks 46 and to the congestion level found in the target queue. This process also performs the admission control. Then, the cell is forwarded to other enqueue unit 22 when the congestion level is not reached and the policy parameters are respected or to Xmit Forwarding Information unit 20 for discarding when the

congestion level is reached or one of the policy parameters has not been respected.

Then, Enqueue unit 22 enqueues the cell in the appropriate queue of queuing area 24 using the cell address pointer. Note that queuing Area 24 which includes essentially the apparatus according to the invention holds the celi buffer address pointers which have been queued until a dequeue occurs.

The Traffic Management Scheduler 26 selects the queue from which a cell has to be transmitted, according to the queue priorities (different qualities of service are provided). The selected queue is then sent to the dequeue unit 28. Scheduler 26 is activated by a status indicator provided by queuing area 24 (at least one cell in the queue).

Dequeue unit 28 dequeues a cell from the scheduled queue, provides the forwarding information found in Control Blocks 16 (destination port, label out). Then, the cell is sent to Xmit Forwarding Information unit 20 which feeds the cells to the CAD module for transmission or discarding.

Control unit 14 (Guided cells) performs all the actions requested by the Control Point, read or write in Control Blocks 16 and hardware registers. Guided cells which are used for reading are the "guided read cells" (GRC) while the guided cells which are used for writing are "guided write cells" (GWC).

Note that Control Blocks 16 reside in the storage of the CAP module (external or internal) and contain different types of control blocks such as lookup resolution tables, connection parameters, queue parameters... The Control Blocks are set/updated by the GWCs from control unit 14, but the different processing units of the CAP module can also update the Control Blocks.

For the implementation of the invention, the control block structure in combination with the cell data flow are illustrated in Fig. 2. First, a plurality of buffer control blocks (BCB) 30 are used to store pointers to the addresses of cells which are stored in data buffer 32. Each BCB includes two fields, a next buffer address used to chain the BCBs to form cell queues and the lock bit L which is set to 0 only if the cell is the last cell of a packet and set to 1 otherwise. In fact, for each cell, only the payload is stored in data buffer 32 while the ATM label is resolved by the lookup unit 12.

The output of lookup unit 12 is a pointer to a Connection Control Block (CCB) 34 which contains the following :

- a REAS-CX bit indicating if the reassembly of cells for VC merging is active for the connection.
- the address of a Reassembly Queue Control Block (RQCB) which is used to assemble the packets of the connection when the reassembly of cells for VC merging is active.
- the address of a Scheduled Queue Control Block (SQCB) assigned to a specific class of service.
- a discard flag bit DFLAG indicating that the packet in progress is being discarded.
- the QoS field used by the traffic management and congestion unit.

There are also Queue Control Blocks 36 which can be Reassembly Queue Control Blocks (RQCBs) which are each used for queuing BCBs of the different connections to be VC merged, or Scheduled Queuing Control Blocks (SQCBs) which are each used for queuing BCBs from RQCBs belonging to packets of a same of class of service.

A Queue Control Block contains the following

- a HEAD field pointing to the BCB of the first cell of a queue corresponding to a same packet.
- a TAIL field pointing to the BCB of the last cell of a queue corresponding to a same packet.
- a CNT field indicating the number of cells in the queue.
- a TH field which indicates either the threshold or maximum size of the reasembled packet in case of RQCB, or the threshold to be processed by the traffic management and congestion unit in case of a SQCB.
- a reassembly bit REAS indicating whether the queue corresponds to a RQCB or a SQCB.
- an aging bit AGING when the queue corresponds to RQCB.

The mechanism of the apparatus according to the invention is illustrated in Fig. 3 which represents four RQCBs 40, 42, 44 and 46 in the upper part and two SQCBs 48 and 50 in the lowest part. There is one RQCB per Connection to be merged in a single VC.

Assuming that the connections queued in RQCB 40 and RQCB 42 belong to the same class of service associated with the same class of service, they are each transferred to the same SQCB when the last cell of a packet partially queued (except the last cell) in each RQCB is received. Thus, transfer unit 52 transfers the contents of RQCB 42 as soon as the lock bit of the BCB corresponding to the received cell is detected as being 0. It must be noted that the BCB corresponding to this last cell is directly queued within the SQCB after the BCBs which are transferred from the RQCB without being queued into the latter.

As illustrated in Fig. 3, the transmission of the contents of the SQCBs is managed by a merged VC scheduler 54. Although various criteria can be considered to determine the scheduling rule used by scheduler 54, an absolute priority from the

highest to the lowest priority of the classes of service is preferably used.

When scheduler 54 has selected a SQCB, it becomes locked by locking unit 56 and dequeues cells contained in the SQCB until it detects the last cell of the packet the lock bit L of which is set to 0. This BCB lock bit L 58 triggers locking unit 56 which unlocks scheduler 54 which runs again to select a new SQCB according to its scheduling rule. Such a lock mechanism insures that the cell stream out of scheduler 54 has no packet interleaving while several SQCBs are to be merged in a single VC.

The flow chart of the steps performed by the apparatus according to the invention is illustrated in Fig. 4. When a cell is received, the discard flag in the connection control block is considered (step 60). If the bit is active, it is checked whether the cell is the first cell (step 62). Note that a first cell of a new packet in AAL5 is derived from the fact that the previous cell of the connection was a last cell as indicated by a bit in the ATM cell header; this information "previous cell was last" being kept in memory in the connection control block. If the cell is not the first cell of a new packet, the cell is discarded (step 64).

If the discard flag was active and the cell is a first cell of a new packet, the discard flag is reset (step 66) and the cell is processed in the same way as if the flag was not active. The reassembly bit in the connection control block is checked (step 68). If it is not active, no reassembly is processed (step 70) ; If it is active, the traffic management and congestion management processes are activated to determine either the cell can be received or must be discarded. Various congestion algorithms may be used (step 72). These algorithms have the threshold and the queue count of the SQCB indicated in the connection control block as inputs as well as the count of

remaining free buffers. The output of the processes are : EPD or early packet discard condition, PPD or partial packet discard condition or no discard condition. In the case EPD, if the cell is the first cell of a new packet, the discard flag in the connection control block is set (step 74) and the cell is discarded (step 64). If the cell is not a fist cell, the cell is received (OK). In the case PPD meaning that the congestion is more severe, the already assembled cells in the RQCB are discarded (the queue is purged) (step 76), the discard flag is set (step 74) and the cell is discarded (step 64). Note that EPD discard condition corresponding to more severe congestion state than EPD one, EPD condition is always met when PPD one is met.

Note that splitting the traffic to be merged in several SQCBs corresponding to various classes of service allows upon cell reception to provide QoS differentiation since the inputs of the congestion algorithm are different. For instance, in case of two classes of service, if the highest priority traffic fills all the bandwidth available for the VC, the count of the lowest priority SQCB will increase until all incoming cells of the lowest priority are discarded. This shows that high priority traffic is not impacted by low priority one.

In case the cell may be received with regard to congestion, the count of assembled cell in the RQCB is compared to the RQCB threshold (step 78). If the threshold corresponding to the maximum allowed size of a packet is exceeded, the already assembled cells in the RQCB are discarded (step 76), the discard flag is set (step 74) and the cell is discarded (step 64). In the case the cell may be received with regard to congestion and to max packet size, the aging bit in the RQCB is reset (step 80).

Then, a check is made to determine whether the cell is the last cell (step 82). If the cell is a not the last cell of the packet, the cell is enqueued in the RQCB (step 84). This is

done by chaining the current tail of the RQCB to the buffer of the received cell using the next buffer address field in the BCB of the current tail (updating the RQCB count and updating the RQCB tail).

In case the cell is the last cell of the packet and it is determined that it is not also the first cell (step 86), the already assembled cells in the RQCB and the received last cell are enqueued in a single operation in the SQCB contained in the connection control block (step 88). For this, 1) the tail of the SQCB is chained to the head of the partially assembled packet in RQCB, 2) the tail of the partially assembled packet in RQCB is chained to the new buffer address, 3) the count in the SQCB is incremented with the cell count of the partially assembled packet in RQCB plus one, and 4) the tail of the SQCB is updated. If the cell is also the first cell of a packet, it is enqueued in the SQCB (step 89) (the tail of the SQCB is chained to the new buffer address and the count in the SQCB is incremented by one).

The last operation (step 88 or 89) is to be compared to a classical operation consisting in first queueing the last cell of the packet in the RQCB, second moving the totally assembled packet in the SQCB, and therefore spends more time resulting from the following steps:

- chaining the current tail of the RQCB to the buffer of the received cell using the NBA information field in the BCB of the current tail.
- updating the RQCB count.
- updating the RQCB tail.
- the tail of the SQCB is chained to the head of the totally assembled packet in RQCB.
- the count in the SQCB is incremented with the cell count of the the partially assembled packet in RQCB.
- the tail of the SQCB is updated.

An aging mechanism is provided as a feature of the enqueuing apparatus according to the invention. The object is to detect that a connection which has a packet being assembled becomes inactive. In such a case, the buffer locations containing the cells pointed by the RQCB must be released to the free buffer queue. Such a mechanism scans periodically all the queues (RQCBs). The period is programmable in a wait timer, the default value being one second.

The implementation of the aging mechanism is represented by the flow chart of Fig. 5. At the beginning there is initialization of n=o (step 90). The first step is to read the queue n (step 92). It is checked whether this reassembly process is active, that is whether the queue is RQCB and REAS bit is set to 1 in the RQCB (step 94). If so, it is determined whether AGING bit is set to 1 in the RQCB (step 96). If not, the AGING bit is set to 1 (step 98). If the AGING was already set to 1, it is checked whether the RQCB is empty (step 100). If not, this means that no cell has been received on the connection since the previous time at which the RQCB was processed since any cell received in the queue resets the AGING bit to 0 (see fig. 4). This is interpreted as a failure. Therefore, the RQCB is purged (step 102) and the AGING bit is reset (step 104).

In case the bit REAS of the RQCB is not set to 1, or the AGING bit is not set to 1, or the RQCB is empty, or the RQCB is not empty and the AGING bit has been reset, the process waits for the wrapping of the wait timer (step 106) and n is incremented to n+1 (step 108) before looping back to the beginning of the process at step 92 of reading the queue.

CLAIMS

1. Apparatus for enqueuing cells belonging to a plurality of connections corresponding to several classes of service in a first switching node of an Asynchronous Transfer Mode (ATM) network wherein at least a set of connections of a same class of service are merged in a single virtual circuit (VC) connecting said first switching node to a second switching node, and comprising a data buffer (32) for storing the cells received by said switching node, a queuing area (24) composed of at least a set of queues associated respectively with the connections of said set of connections and a scheduled queue corresponding to said class of service, and wherein the contents of said set of queues are transferred into said scheduled queue before being transmitted on said virtual circuit ;

 said apparatus being characterized in that each queue of said set of queues is associated with a Reassembly Queue Control Block (RQCB) (40, 42, 44 or 46) defining a chain of Buffer Control Blocks (BCB) (30) wherein each BCB corresponds to a cell belonging to a packet transmitted in the associated connection and is composed of the next buffer address in said data buffer and a lock bit (BCB lock) which is normally set to 1 and set to 0 only if said cell is the last cell of said packet, said chain of BCBs being transferred to a Scheduled Queue Control Block (SQCB) (48 or 50) associated with said scheduled queue when it is detected that the lock bit of the cell being stored in said data buffer is set to 0, the corresponding BCB being chained to said chain of BCBs in said SQCB without being previously queued in said RQCB.

2. Apparatus according to claim 1, wherein said plurality of connections includes several sets of connections wherein each set of connections is associated with a class of service, and further comprising a merged VC scheduler (54) for scheduling the transmission of cells from the scheduled queues respectively associated with the different SQCBs (48, 50) and corresponding respectively to the different classes of service.

3. Apparatus according to claim 2, wherein said merged VC scheduler (54) performs the scheduling of cell transmission based upon a priority order of said classes of service, and comprising locking means (56) for locking the transmission of cells from a scheduled queue associated with a SQCB until the end of the entire packet which the cells belong to.

4. Apparatus according to claim 3, wherein said locking means (56) unlocks said merged VC scheduler (54) when the lock bit (58) of the BCB corresponding to the cell to be transmitted is set to 0 indicating that said cell is the last cell of said packet.

5. Apparatus according to any one of claims 1 to 4 containing an aging mechanism which is periodically activated for deleting the cells already enqueued in a queue associated with a RQCB when no cell has been enqueued in said RQCB during a predetermined period of time.

6. Apparatus according to claim 5, wherein said RQCB contains an aging bit (AGING), said aging bit being automatically set to 0 when a new cell is enqueued in said RQCB, and said aging mechanism being periodically activated for deleting the cells enqueued in said RQCB when said aging bit is set to 1 and for setting said aging bit to 1 if it is equal to 0.

7. Apparatus according to any one of claims 1 to 6, wherein a Connection Control Block (CCB) (34) is associated with each connection to be merged, said CCB including the address of the RQCB (40, 42, 44, 46) used to assemble the packets corresponding to said connection, the address of the SQCB (48, 50) used to enqueue the cells corresponding to said RQCB, and a flag bit (DFLAG) indicating that the packet in progress is being discarded.

8. Apparatus according to claim 7, further comprising an Early Packet Discard (EPD) means for setting said DFLAG bit in said CCB (34) when activated and for discarding the cells of the packet being received if said means is activated before receiving the first cell of said packet.

9. Apparatus according to claim 7, further comprising a Partial Packet Discard (PPD) means for setting said DFLAG bit in said CCB (34) when activated, and purging said RQCB by deleting the cells of a packet already queued and discarding the following cells of said packet.

ENQUEUING APPARATUS FOR VIRTUAL CIRCUIT MERGING IN A ROUTER OF AN ASYNCHRONOUS TRANSFER MODE (ATM) NETWORK

Abstract

Apparatus for enqueuing cells belonging to a plurality of connections in a switching node of an Asynchronous Transfer Mode (ATM) network wherein at least a set of connections of a same class of service are merged in a single virtual circuit (VC), and comprising a set of queues associated respectively with the set of connections and a scheduled queue corresponding to the class of service, and wherein the contents of the set of queues are transferred into the scheduled queue before being transmitted on the virtual circuit. Each queue of the set of queues is associated with a Reassembly Queue Control Block (RQCB) (40, 42, 44 or 46) defining a chain of Buffer Control Blocks (BCBs) wherein each BCB corresponds to a cell belonging to a packet transmitted in the associated connection and is composed of the next buffer address and a lock bit (BCB lock) which is normally set to 1 and set to 0 only if the cell is the last cell of the packet, the chain of BCBs being transferred to a Scheduled Queue Control Block (SQCB) (48 or 50) when the lock bit of the cell is set to 0, the corresponding BCB being chained to the chain of BCBs in the SQCB without being previously queued in the RQCB.

FIG. 3

FR 9 99 077
Aznar et al
1/5

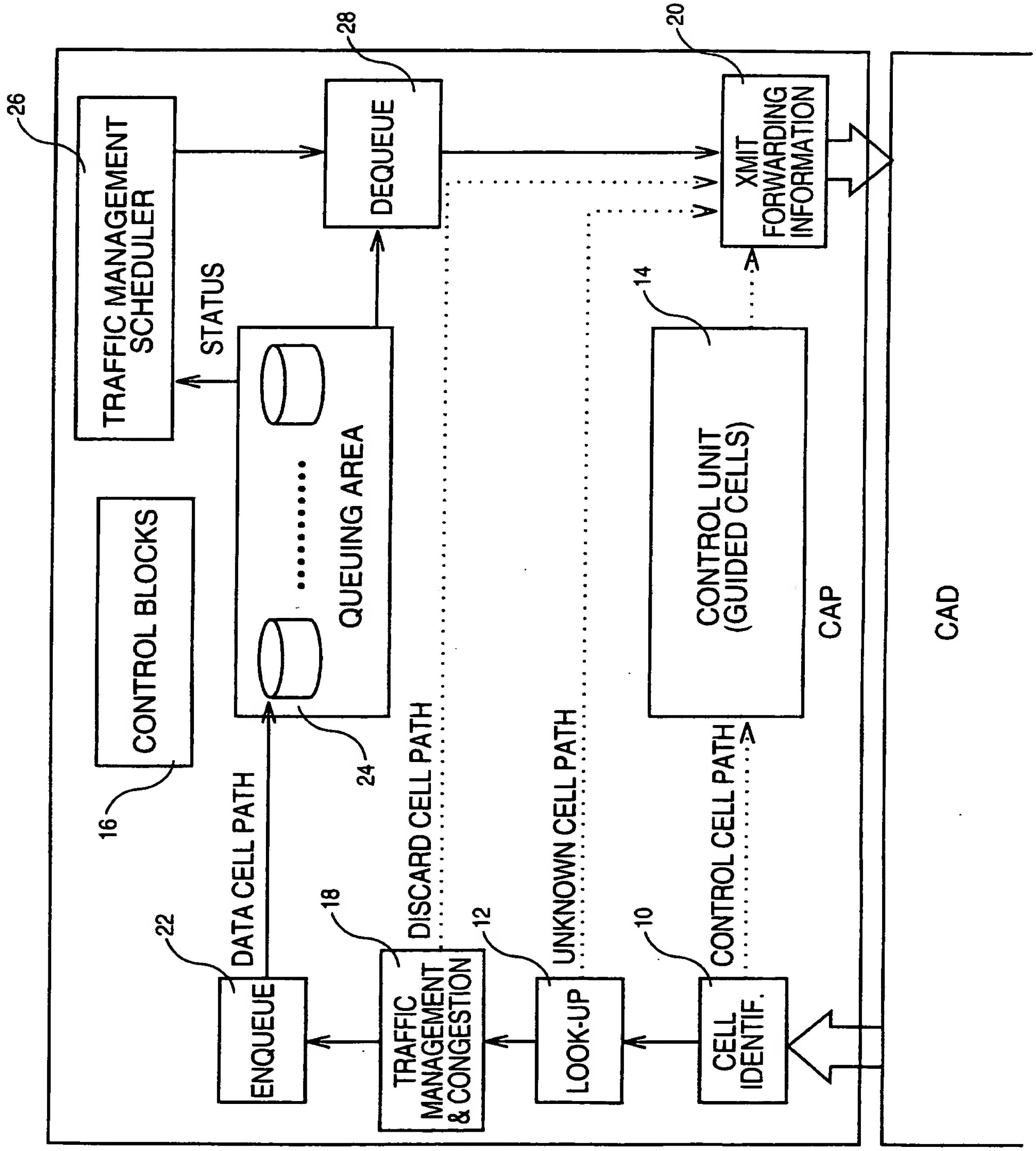

FIG. 1

FR 9 99 077
Aznar et al
2/5

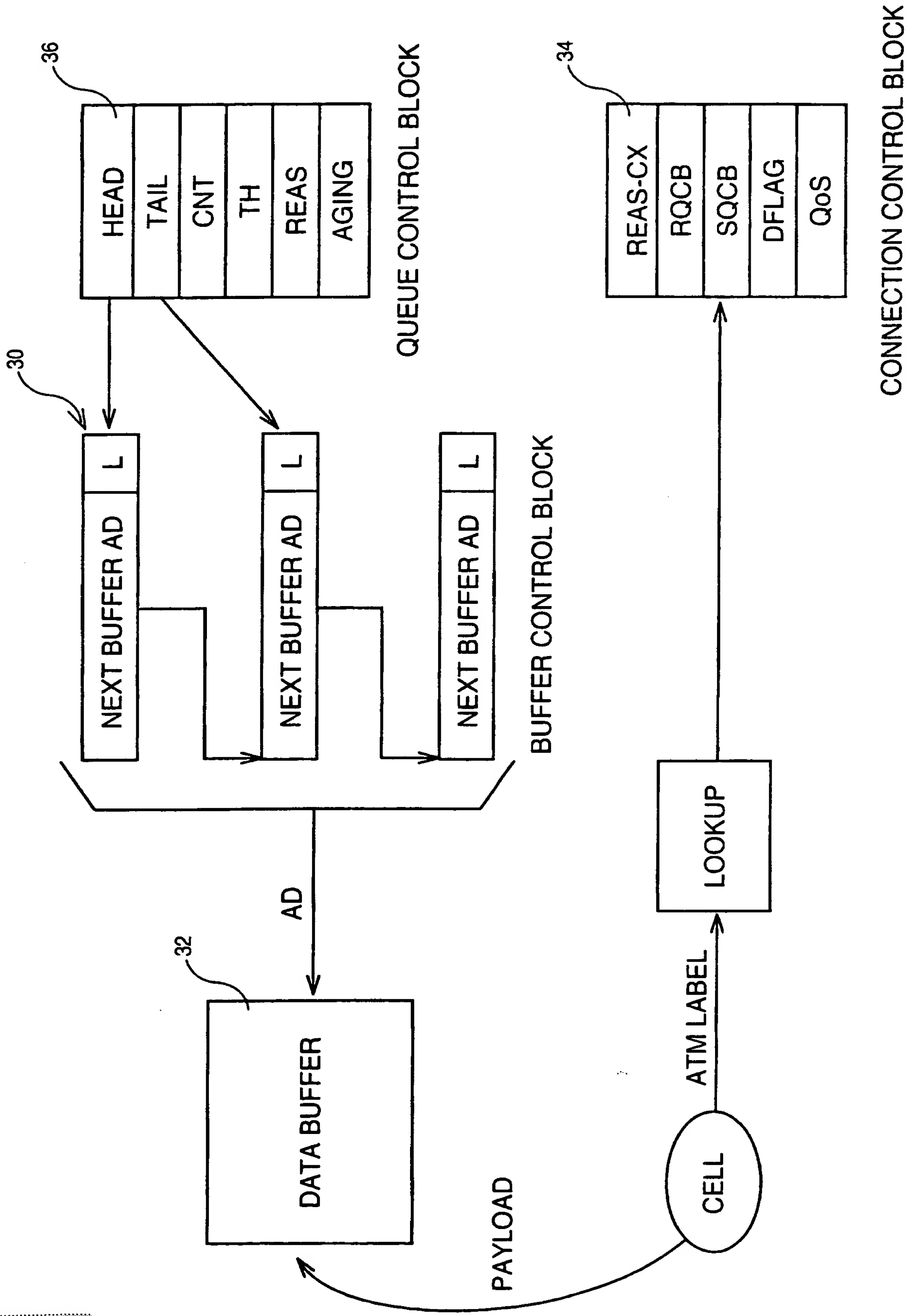

FIG. 2

FR 9 99 077
Aznar et al
3/5

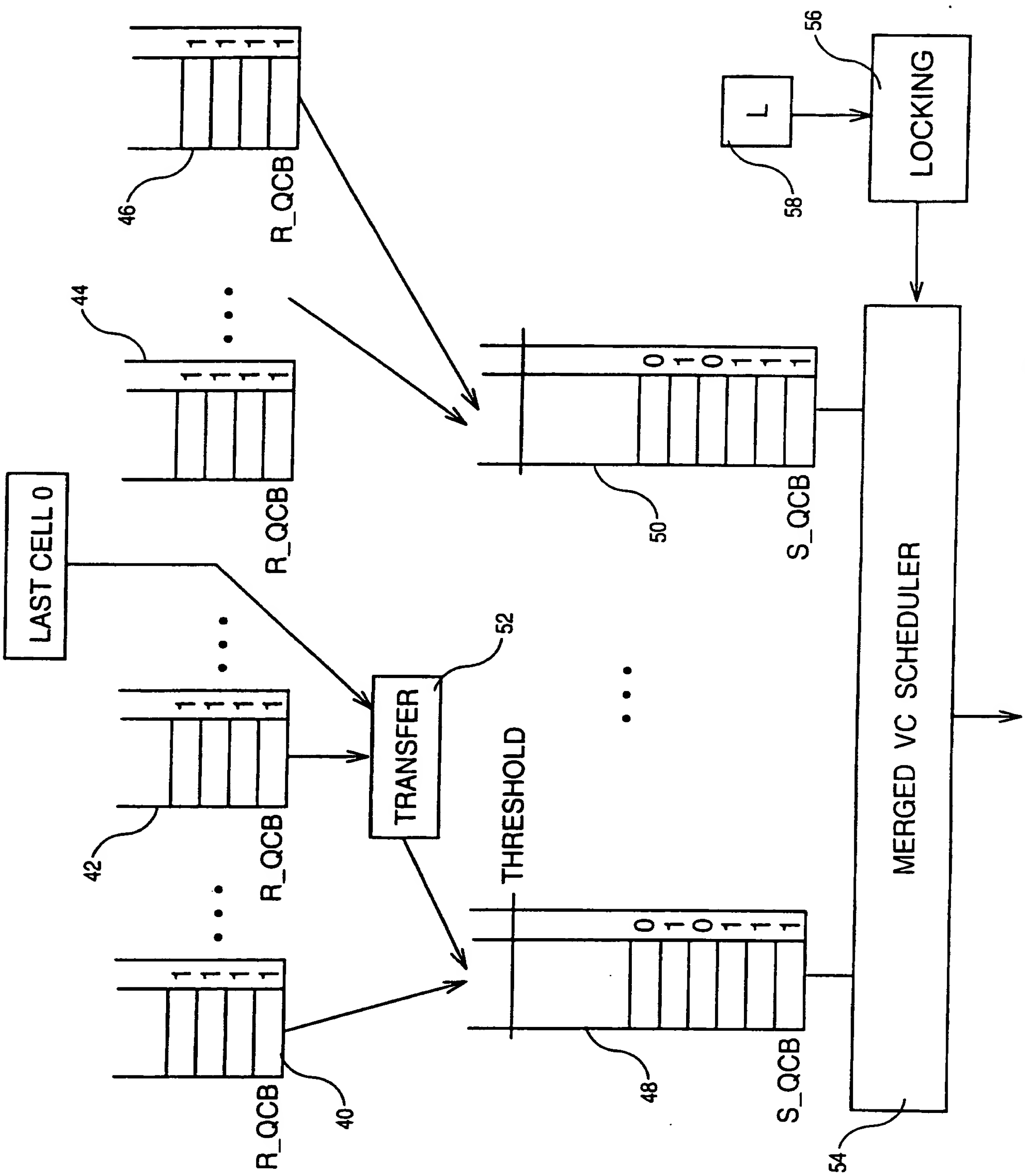

FIG. 3

FR 9 99 077
Aznar et al
4/5

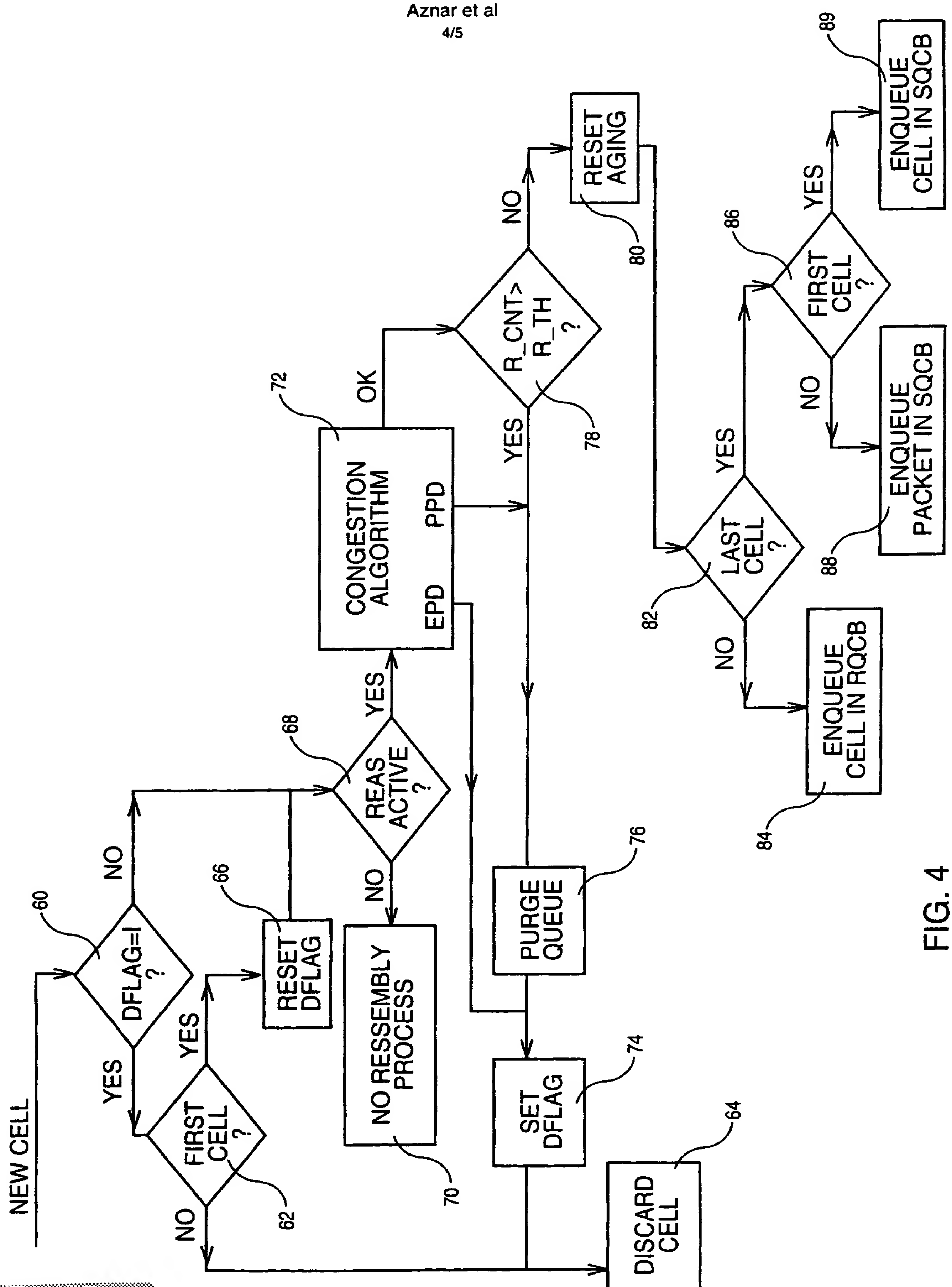

FIG. 4

FR 9 99 077
Aznar et al
5/5

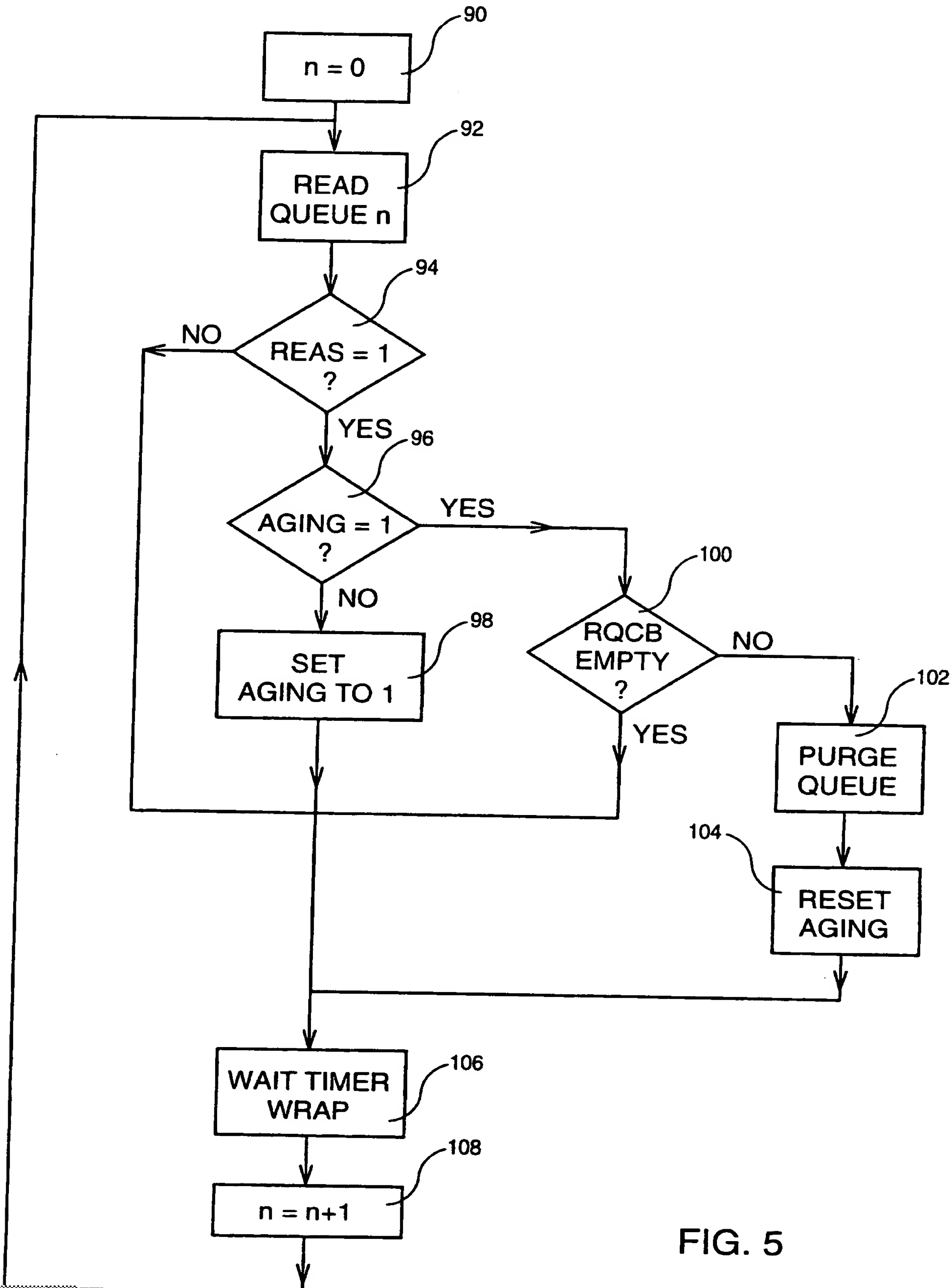

FIG. 5